Prosody-based Spoken Algerian Arabic Dialect Identification

Soumia BOUGRINE*, Hadda CHERROUN*, Djelloul ZIADI†
*Laboratoire d'informatique et Mathématiques
Université Amar Telidji Laghouat, Algérie
{sm.bougrine, hadda_cherroun}@lagh-univ.dz
†Laboratoire LITIS - Université Normandie Rouen, France
Djelloul.Ziadi@univ-rouen.fr

***Abstract*—Dialect is the most common way of communication in every day life. Automatically identifying a dialect is a challenging task, especially when we deal with close dialects within the same country. In this paper, we address the problem of Spoken Algerian Arabic Dialect Identification (SAADID). Indeed, we propose a new system based on prosodic speech information, namely intonation and rhythm. The rhythm features are got using a coarse-grained consonant/vowel segmentation. The performance of this approach is shown through experiments on six dialects of the departments of Adrar, Algiers, Bousaâda, Djelfa, Laghouat and Oran. The results prove the suitability of our prosody-based system for SAADID with more than 69% of precision using 2s test utterances.**

Keywords—Algerian Arabic Dialect, Dialect Identification, Prosodic Modeling, Syllable' Nuclei Segmentation.

I. INTRODUCTION

Dialect identification (DID) is a part of Natural Language Processing. It is the task of automatically recognizing a dialect from a sequence of speech in a known language [1]. It is more complex than automatic language identification due to the utilisation of the same language. Spoken dialect identification techniques are used in many fields such as detection/classification for a Spoken Document Retrieval System (SDRS), for incrementing the performance of automatic speech recognition (ASR), and speaker recognition or multi-language translation systems.

Arabic is a semitic language, spoken by more than 420 million people in 60 countries in the world [2]. It has multiple variants: Ancient Arabic (AA), Classical Arabic (CA), Modern Standard Arabic (MSA), and Dialectal Arabic (DA) [3]. AA is no longer used, we find it in the old writing literary essentially the poems. CA is the language of the Coran, which is the principal source of grammar and phonetic rules. It was spoken in the seventh century. MSA is the official language of all Arab countries; used in administrations, schools in modern literature, official radios, press, some TV programs, etc. DA is often referred to colloquial Arabic, or vernaculars; which is used in public places and situations of informal communications. There is a large number of dialects that can be grouped in five categories: Arabian Peninsula, Levantine, Mesopotamian, Egyptian and Maghrebi [4].

Algerian arabic dialect is a maghrebian dialect. It has a many variations which are mainly developed as a result of both phases of arabization and deep colonization history. According to the arabization phases, we can divide dialects into three major groups: pre-Hilālī, Hilālī and the mixed dialects. The pre-Hilālī are called rural, sedentary dialects; which are spoken in area that is affected by the expansion of Islam in the 7th century. The Hilālī dialects take name from Banū Hilāl, it is named Bedouin dialects; spoken in area which is influenced by the Arab immigration in the 11th century. The mixed dialects or urban Hilāl are spoken in region that is affected by both arabization phases [5], [6]. Furthermore, Algerian dialects are also influenced by the long period of a deep colonization. In fact, Arabic Algerian dialect is affected by other languages such as Turkish, French, Italian, and Spanish [7].

A language can be distinguished from an other by means of many knowledges extracted from the speech information. Indeed, we can use acoustic-phonetic (spectrum, phone inventory), phonotactic (sequence of sounds), prosodic (duration, pitch, intonation), lexical (vocabulary and morphology word) and syntactic (phrases and grammar of syntax); which are from low-level to high-level, the features of the speech [8]. The phonemes are the smallest segments of speech sound in a language, where their number and frequency are different from language to another. Phonotactic information are the features of concatenated phonemes, whereas, prosodic ones are extracted from the study of three phenomena: accentuation, intonation and rhythm.

Any system of language/dialect processing is mainly composed of three steps: segmentation, feature extraction and modeling. The first and second steps are called the front-end while the last step is called the back-end of the system. Concerning the dialect identification systems, there is a lot of interest for non semitic languages. However, little attention has been paid to the Arabic Dialect Identification (ADID) when the input is a speech compared with written text. Specially when dialects are very close, we mean dialects from the same country. Indeed, some tentatives have only been carried about ADID inter-countries by considering different countries [9], [1], [10], [11]. Since, few of them have studied the dialects of the same country [12]. Recently, a number of works have attempted to perform written ADID [13], [14], [15].

In addition, many works have focused on the exploitation of the speech information at the acoustic and phonotactic levels [9], [1], [10]. In contrast, few tentatives have tried to use prosodic knowledge [11], [12]. However, it has been proved that dialect variations are pursued in phonetic and prosodic features of speakers [16].

In this paper, we propose an ADID for spoken Algerian dialects based on prosodic variations. The rest of this paper is organized as follows. In the next section, we review some related works on DID approaches and we focus on those ones dedicated to ADID. We describe and explain our approach in Section III. Experiments and results are described and commented in Section IV.

II. Related Works

As previously mentioned, acoustic-phonetic, phonotactic and prosody levels are exploited alone or combined to identify dialects in general. As far as we know, the pioneer work has begun in the middle of the nineties and is due to [17]. Before that, let us mention that for Arabic dialects and especially for Algerian ones there are a noticeable lack of speech databases dedicated to scientific researches purposes. Furthermore, there is even less standard ones.

Let us first review some works where acoustic-phonetic models are used alone. Most of them are based on spectral features, which is Mel Frequency Cepstral Coefficient (MFCC) or Shifted Delta Cepstrum (SDC), with Gaussian Mixture Model (GMM) and Vector Space Modeling (VSM) [18], [9], [19]. For instance, Torres-Carrasquillo et al. [20] have used GMM with SDC characterization using CallFriend and Miami corpora from three languages: English, Chinese and Spanish. They selected two dialects for each language. Separately, they developed dialect identification system for each language. The best achievement between this two corpus is an average of 13% Equal Error Rate (EER) for CallFriend corpus.

Specifically for Arabic dialects, Alorfi [18] presents different acoustic-phonetic approaches by using Hidden Markov Models. He associated two states for each dialect representing common and unique sounds respectively. He applied this approach on only one Egyptian dialect and one Gulf dialect. The best result of this system is more than 96% of correct identification.

Biadsy et al. [21] used a phone recognizer to label phones and built on this segmentation a GMM-supervectors. However, this approach constrained the acoustic model. This model with an SVM classifier using special Kernel function is applied on four Arabic dialects (Iraqi, Gulf, Levantine, Egyptian). This approach achieved an EER of 6.1%. Al-Ayyoub et al. [9] designed an acoustic model using fixed size segmentation for which they extracted the features using Wavelet transform with a significant feature reduction. They deal with two dialects Jordanian and Egyptian. They achieved 97% precision barrier. However, as they confirmed, this result is not conclusive due to the limited size and the quality of the database. For the context of Magrebian dialects automatic identification, Lachachi and Adla [19] instrumented the Universal Background Model that they reduce to support special SVM classification. Indeed, they reduce the size of database by using the Minimal Enclosing Ball method by means of a fuzzy C-mean clustering algorithm. They deal with a database containing five regions (Oran, Algiers, Constantine, Morocco and Tunisia). The best performance is achieved with a precision rate of 80.49%.

Elsewhere in phonotactic cues, there are few attempts that have tried to build ADID by using Parallel Phone Recognition followed by Language Modeling (PPRLM) [1], [22]. For instance, Biadsy et al. [1] have applied a PPRLM using nine (Arabic and non-Arabic) phone recognizers. They performed their experiments on a large database of four Arabic dialects (Egyptian, Gulf, Iraqi, Levantine) plus MSA. Their system accuracy for 30s test utterances is more than 81%. In another work, Akbacak et al. [22] have designed an approach that combines between three models to identify four Arabic dialects (Iraqi, Gulf, Levantine and Egyptian). These models are a cepstral GMM, a PPRLM and Phone Recognition modeled via SVMs (PRSVM). The combination is performed at the score-level. The best result of this system is 98% of correct identification for pairwise classification. More interesting, some works have exploited both acoustic-phonetic and phonotactic features to perform ADID [23], [24], [25]. Indeed, Greenberg et al. [23] have designed a combined approach, using four core classifiers based on three spectral similarities and n-gram. The combination of four classifiers is done at the back-end level of the system using Bayes classifier. They target a set of 24 languages containing four variations of Arabic language, which are MSA and three dialects: Iraqi, Levantine and Maghrebi.

In a more high layer of the speech knowledge abstraction, in this case the "prosody", we have observed that there are very few attempts. Prosodic information can be picked up from three observed phenomena: accentuation, intonation and rhythm. Recently for accent identification of Swiss French, Lazaridis et al. [26] have developed a system that uses a syllabic segmentation, from which extracts features that expressed main aspects of prosody such as duration, intensity and pitch. Their classifier is an SVM-based method with a moderate corpus which deals with four Swiss French accents (Geneva, Martigny, Neuchâtel and Nyon). The accuracy rate of this system is 45.72% where the number of syllables is equal or bigger than 20 in test utterances.

Based on a previous work [27] which claims that some Arabic dialects (Syria, Jordan, Morocco, Algeria, Tunisia and Egypt) can be regrouped on three areas using rhythmic information, Rouas et al. [11] have designed a system of semi identification for three dialectal areas: Maghreb (Morocco, Algeria), Middle-East (Syria, Jordan), and an intermediate one (Tunisia, Egypt). Their approach gathered all prosodic information: accentuation, intonation and rhythm. Indeed, they use a segmentation based on consonant/vowel location to get the approximative structure of the syllable. They have used a special codification to represent the duration of phonemes and the energy instead of real values. A fundamental frequency (F_0) is associated with a syllable. Then, they classified dialects using multi-gram models, where grams are their pseudo-syllables and each area's dialect is represented by the most frequent sequences of n-gram. Unfortunately, they tested their system on a small database. This system reached 98% of correct identification of the three dialectal areas. An other work on Arabic dialect identification has been proposed by Biadsy et al. [28] which combined the prosodic and phonotactic approaches. In fact, they augmented their phonotactic system, above reviewed, by adding some prosodic features like duration, F_0. They tested this system on four Arabic dialects (Gulf, Iraqi, Levantine, and Egyptian). The identification accuracy of this system is 86.33% for 2 minutes test utterances.

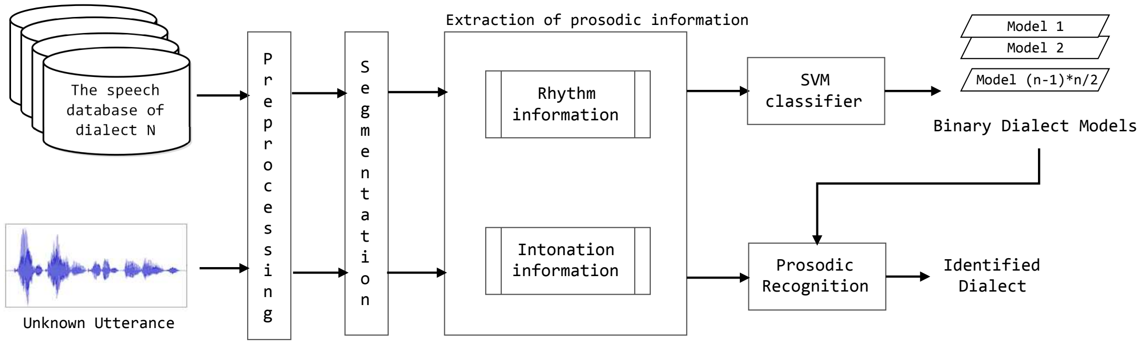

Fig. 1. General view of our dialect identification system.

III. Our Dialect Identification Approach

In this preliminary work, we aim to design an identification system for Algerian Arabic dialects. Indeed, in order to identify a dialect from a speech, we rely on machine learning techniques. For this task, we need to characterize the speech signal by some features after its segmentation. So, we need to fix the level from which we extract such features. For this purpose, let us recall the knowledge that we have, up today, in Arabic dialects and specially in Algerian ones:

- The classification of Arabic dialects depends on prosodic phenomenon. Indeed, Barkat et al. [29] have shown that the prosodic information can be sufficient to identify Arabic dialects.
- Arabic dialects present significant differences at the syllable structure [30], which can be emerged by the mensuration of rhythm metrics [31].
- Intonation information salient the discrimination between Arabic dialects [32], [33].
- Specially for Arabic Algerian dialects, Benali claims that rhythm and intonation are very discriminant parameters [12].

On the light of these information, we focus in our approach on the use of prosodic features: rhythm and intonation to detect the dialect by using an approximative consonant/vowel segmentation to capture prosodic measures. One question that needs to be resolved is that which segmentation we have to perform? The best alternative is to perform phoneme segmentation (consonant/vowel). However, for Arabic the problem of phoneme segmentation is not kindly resolved and unfortunately there is no decoder for Arabic dialect such as the case for other dialects. In order to circumvent this problem, we perform a coarse-grained C/V segmentation as it will be explained below.

Let first remind that the syllable is the unit of pronunciation between phoneme and word. It is divided into three components: the opening *Onset*, the central *Nucleus*, the closing segment *Coda*. The nucleus and coda are called the rhyme (or rime) [34]. In Arabic dialect, each syllable contains at least a nucleus while coda and onset are optional. The nucleus is imperatively one or many vowels and the others are consonants. One can model a syllable structure by the regular expression $C^*V^+C^*$ where C is a consonant and V is a vowel.

Let us back to the segmentation. As in rhythm we need just to measure for consonant and vowel segments their ratios and metric duration, thus, we sketched up these measures by considering nuclei as vowel segments and what remains as consonant segments. It is what we call coarse-grained segmentation. For example, consider the sequence of the three syllables presented in Figure 2 where syllable segmentation provides (b) and we get our coarse-grained segmentation (c) where *VS* (*CS*) represents vowel (consonant) segment.

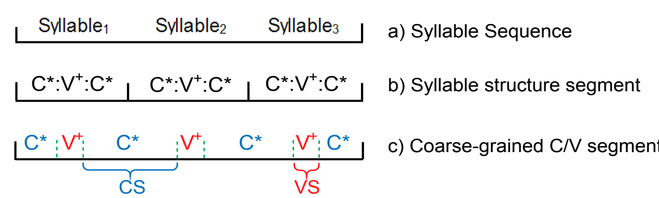

Fig. 2. Coarse-grained segmentation.

Our approach is illustrated by Figure 1. Given a database of speech from different dialects, within the preprocessing, we remove noise and silence from all speeches of the database. After that a syllable segmentation is performed. In our framework, this is done automatically by using Praat tool [35] enhanced by the script Prosogram v2.9 [36]. The latter performs syllable segmentation by using intensity of band-pass filtered signal. The syllable' nuclei are delimited using spectral and amplitude changes.

As mentioned above, among the prosodic information, we use intonation and rhythm. In fact, our speech is characterized by a vector composed of a pair of feature sets. In the first set, we consider three groups of intonation metrics for each sentence:

1) Global information or Pitch Range: four pitch values of nucleus are calculated using quantiles, such as *Bottom* (2%), *Median* (50%), *Top* (98%) in Hz and *Range* is the difference between *Top* and *Bottom* in SemiTones (ST).
2) Tone direction information: has three metrics based on the proportion (%) of pitch direction of syllable nuclei with *Static level* (Gliss), *Rising* and *Falling* tones.
3) Total size pitch trajectory: this group has two metrics calculated using pitch trajectory (sum of absolute intervals) within/between syllabic nuclei divided by duration (in ST/s).

TABLE I. DESCRIPTION INTONATION AND RHYTHM METRICS.

Parameters	Feature	Definition
F0 Global	Bottom	2 th quantiles pitch nucleus
	Top	98 th quantiles pitch nucleus
	Median	50 th quantiles pitch nucleus
	Range	Difference between Top and Bottom
F0 Tones	Gliss	The proportion of static level tones
	Rises	The proportion of dynamic rising tones
	Falls	The proportion of dynamic falling tones
F0 movement	TrajIntra	Pitch trajectory of nuclei / duration
	TrajInter	Pitch trajectory between nuclei / duration
Ramus Model	%V	The proportion of vocalic interval
	ΔV	The standard deviation of vocalic interval
	ΔC	The standard deviation of consonantal interval
Dellwo Model	varcoV	ΔV / mean of vocalic interval duration
	varcoC	ΔC / mean of consonantal interval duration
Grabe Model	rPVI-V	Pairwise Variability Indice of vocalic interval
	rPVI-C	Pairwise Variability Indice of consonantal interval
	nPVI-V	Normalized PVI of consonantal intervals
	nPVI-C	Normalized PVI of vocalic intervals
Other	MeanC	Mean of consonantal intervals
	MeanV	Mean of vocalic intervals
	Speech Rate	Syllable par second

In the second feature set, we consider rhythm information by using three models; *Ramus* [37], *Grabe* [38] and *Dellwo* [39]. In addition, we consider two other metrics: *Speech Rate* and *Mean* duration. In order to compute the metrics of these models, we use the segmentation system as mentioned before where the nuclei segments are considered as vowel segments and the others parts –Onset and coda– as consonant segments. From Ramus model, we use three metrics: the duration proportion, the standard deviation of vocalic interval ($\%V$, ΔV), and the standard deviation of consonantal interval ΔC. Their normalization version according to the speech rate is the model of Dellwo, called *varcoV* and *varcoC*. For both intonation and rhythm, metrics are calculated for each sentence. Table I gives more details on these features.

After this front-end step of the system, the back-end phase begins. In fact, in this step, we use SVM binary classification to create our dialect models. Indeed, SVM methods have proven their efficiency as language/task independent methods. Each model is calculated with One-Versus-One technique.

IV. EXPERIMENTS AND RESULTS

In order to measure the performances of our approach, we need a database of speeches. Unfortunately, there are no standard corpus available for Arabic Algerian speeches such as for others countries [40]. For this reason, we have been constrained to prepare our own corpus. This task of collecting a large database is very challenging. As the country is very large and many dialects exist and people are very fair to record their own voices. We have targeted 15 dialects, but due to balancing sizes considerations, we have limited our experiments in this preliminary version to six dialects. These dialects are from the departments of Algiers, Adrar, Bousaâda, Djelfa, Laghouat and Oran. Figure 3 illustrates their geographical distribution.

The deployed corpus for experiments includes two parts of speeches: our Own Recording database (OR) and speeches extracted from reports which are selected from regional Radios and TVs (RTV). The OR speeches includes 34 speakers (1.5h) from 18 different departments. Each speaker produces a set of 57 sentences. These latter have two parts: MSA (10 sentences) and dialect speech (47 sentences) [41]. In this work, dialect speech consists of free responses, free translation of phrases and a short text story 'The North Wind and the Sun' and a semi-guided narration obtained from a series of images without text 'Frog, where are you?' and MSA. The RTV speeches are added, which are extracted from 13 videos and reports (2h).

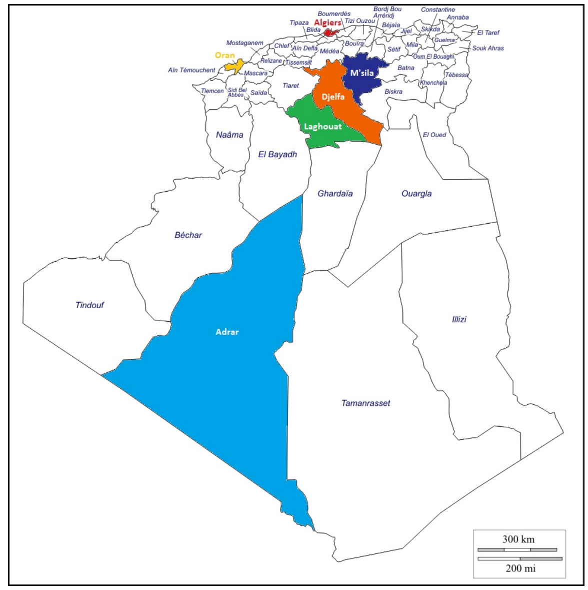

Fig. 3. Geographical distribution of studied dialects.

TABLE II. DETAILS OF THE SPEECH DATABASE.

Dialect of	Classification	#Sentence	#Speaker	Ratio (OR/RTV)
Oran (ORN)	Mixed	126	16	84/42
Algiers (ALG)	Mixed	142	9	61/81
Laghouat (LAG)	Hilālī	161	16	110/51
Adrar (ADR)	Mixed	169	12	96/73
Bousaada (BOU)	Hilālī	170	8	150/20
Djelfa (DJE)	Hilālī	173	3	173/0
Total	/	941	64	674/267

Some preprocessing are performed manually. We have filtered all French portions, stuttering and pauses. In addition, we have discarded sentences that uses many words lent from French. After preprocessing, the duration of utterance is about 2s in average. We have ensured, when it is possible, to have balanced number, for speakers' age and gender. The speakers are adult population. Our database can be considered as moderate in term of time duration but it is balanced in term of number of sentences per dialect. Table II gives some details on this database. The last column reports the ratio between OR and RTV sentences.

The intonation metrics and speech rates are extracted using Prosogram script. This script is also used to perform syllable nuclei segmentation. The outcome of this segmentation gives annotated files. These latter are codified by replacing all nuclei by V and other by C. Then, within these annotated files, the rhythm features are calculated using Correlatore[1] Software which is a program designed for rhythmic analysis. At the end of this steps, each sentence is characterized by a vector

[1]http://www.lfsag.unito.it/correlatore

TABLE III. MATRIX OF CONFUSION GIVEN BY THE PROSODIC MODEL (CORRECT IDENTIFICATION RATE : 69.18 % (651/941)).

	ORN	ALG	LAG	ADR	BOU	DJE
ORN	**68**	16	2	22	6	12
ALG	14	**91**	8	12	10	7
LAG	1	3	**123**	17	9	8
ADR	7	5	5	**135**	9	8
BOU	1	8	17	12	**114**	18
DJE	8	10	9	6	20	**120**

that contains all above metrics.

The dialect models are made using SVM method with Universal Pearson VII Kernel (PUK) function. For the generation of these models and testing parts, we have used the Weka tool [42], which is one of the most commonly used tools in machine learning.

In order, to ensure that our results are reliable, we use the k-fold cross-validation technique with k=10. Figure 4 illustrates from left to right the Precision, the Recall, and F-Measure of the system. Table III reports the matrix of confusion.

First, let us analyze the performances of the whole system. It gives an average of **69.3%** Precision, with **69.2%** Recall and **69%** F-Measure, which are considered as good results because the departments are next to each other and thus the dialects are very close.

Concerning the details, the best results of the system are achieved for dialect of Laghouat where 75% are correctly detected. The worst results are observed for the dialects of Oran then Algiers with F-Measure about 60.4%, 66.2% respectively. This is mainly due to the fact that Oran and Algiers are metropolitan cities and most of their population are not native.

Concerning results of Djelfa dialect identification, one can observe, from Table III, that more than 11.5% of Djelfa sentences are confused with dialect of Bousaâda. The same observation can be noted in the opposite direction. One can argue that this is due of the fact that the departments are geographically very close.

As observed, the identification of dialect using our corpus gives better results comparing with those represented by sentences from RTV. Indeed, we don't have any information about the origin of speakers, whether they are native or not.

V. CONCLUSION

In this paper, we have designed an approach that identifies Arabic Algerian dialects from a speech. Our classification-based approach characterizes a speech at the prosody level. Indeed, intonation and rhythm measurements are used as vector features and dialects learner are constructed using Support Vector Machines method.

As far as we know, in the context of Algerian Arabic dialect identification, this is the first attempt based on prosodic information. Certainly the database has moderate size but this framework can be considered as a kernel for more complete systems. The reached results are very satisfying. We work on extending the database using our own recording. We are also working on multi-classification by a smart weighting of both rhythm and intonation features.

As some features that are also discriminative such as nasalization and allophones. This can be achieved by combining information from the acoustic level and the prosodic one. In some other investigation we are focusing on enhancing our approach by considering accentuation features.

REFERENCES

[1] F. Biadsy, J. Hirschberg, and N. Habash, "Spoken Arabic Dialect Identification Using Phonotactic Modeling," in *EACL*, Athens, Greece, 2009, pp. 53–61.

[2] M. P. Lewis, F. S. Gary, and D. F. Charles. (2015) Ethnologue: Languages of the World, Eighteenth edition. Web. Dallas, Texas: SIL International. [Online]. Available: http://www.ethnologue.com/statistics/size

[3] M. Embarki, "Les dialectes arabes modernes: état et nouvelles perspectives pour la classification géo-sociologique," *Arabica*, vol. 55, pp. 583–604, 2008.

[4] K. Versteegh, *The Arabic Language*. Cambridge: Edinburgh University Press, 1997.

[5] H. Palva, "Dialects: classification," *Encyclopedia of Arabic Language and Linguistics*, vol. 1, pp. 604–613, 2006.

[6] C. Pereira, "Arabic in the North African Region," in *Semitic Languages. An International Handbook*, S. Weniger, G. Khan, M. P. Streck, and J. C. E. Watson, Eds., Berlin, 2011, pp. 944–959.

[7] J. Leclerc. (30 avril 2012) Algérie dans l'aménagement linguistique dans le monde. Web. Québec, TLFQ, Université Laval. [Online]. Available: http://www.tlfq.ulaval.ca/axl/europe/danemark.htm

[8] H. Li, B. Ma, and K. Lee, "Spoken language recognition: From fundamentals to practice," *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1136–1159, 2013.

[9] M. Al-Ayyoub, M. Rihani, N. Dalgamoni, and N. Abdulla, "Spoken Arabic dialects identification: The case of Egyptian and Jordanian dialects," in *5th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan*, 2014, pp. 1–6.

[10] M. Barkat-Defradas, R. Hamdi, and F. Pellegrino, "De la caractérisation linguistique à l'identification automatique des dialectes arabes," in *MIDL*, Paris, France, 2004, pp. 51–56.

[11] J. L. Rouas, M. Barkat-Defradas, F. Pellegrino, and R. Hamdi, "Identification automatique des parlers arabes par la prosodie," in *Journées d'Etude sur la Parole*, Dinard, France, 2006, pp. 193–196.

[12] I. Benali, "Le rôle de la prosodie dans l'identification de deux parlers algériens: l'algérois et l'oranais," in *MIDL*, Paris, France, 2004, pp. 128–132.

[13] M. Al-Badrashiny, H. Elfardy, and M. Diab, "AIDA2: A Hybrid Approach for Token and Sentence Level Dialect Identification in Arabic," in *CoNLL*, Beijing, China, July 2015, pp. 42–51.

[14] F. Sadat, F. Kazemi, and A. Farzindar, "Automatic Identification of Arabic Language Varieties and Dialects in Social Media," in *SocialNLP*, Dublin, Ireland, 2014, pp. 22–27.

[15] O. Zaidan and C. Callison-Burch, "Arabic Dialect Identification," *Computational Linguistics*, vol. 40, no. 1, pp. 171–202, 2014.

[16] J. C. Wells, *Accents of English*. Cambridge University Press, 1982, vol. 2.

[17] M. A. Zissman, T. P. Gleason, D. M. Rekart, and B. Losiewicz, "Automatic dialect identification of extemporaneous conversational, Latin American Spanish speech," in *ICASSP*, Atlanta, Georgia, 1996, pp. 777–780.

[18] F. S. Alorfi, "Automatic Identification of Arabic Dialects Using Hidden Markov Models," Ph.D. dissertation, University of Pittsburgh, USA, 2008.

[19] N. Lachachi and A. Adla, "GMM-Based Maghreb Dialect Identification System," *Journal of Information Processing Systems*, vol. 11, no. 1, pp. 22–38, March 2015.

[20] P. A. Torres-Carrasquillo, T. P. Gleason, and D. A. Reynolds, "Dialect Identification Using Gaussian Mixture Models," in *ODYSSEY*, Toledo, Spain, 2004, pp. 297–300.

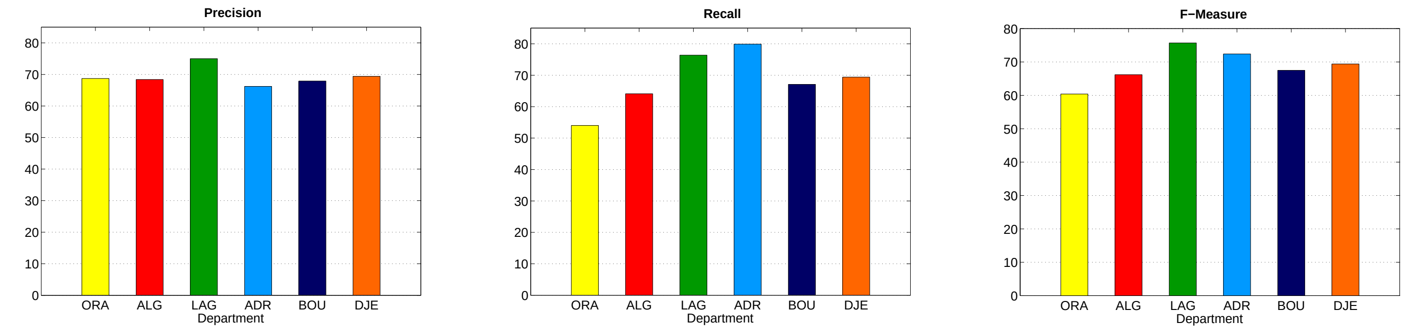

Fig. 4. System Performances.

[21] F. Biadsy, J. Hirschberg, and D. P. W. Ellis, "Dialect and Accent Recognition Using Phonetic-Segmentation Supervectors," in *Interspeech, Florence, Italy*, 2011, pp. 745–748.

[22] M. Akbacak, D. Vergyri, A. Stolcke, N. Scheffer, and A. Mandal, "Effective Arabic Dialect Classification Using Diverse Phonotactic Models," in *Interspeech, Florence, Italy*, 2011, pp. 737–740.

[23] C. Greenberg, A. Martin, and M. Przybocki, "The 2011 NIST Language Recognition Evaluation," in *Interspeech, Portland, Oregon, USA*, 2012.

[24] A. Hanani, M. J. Russell, and M. J. Carey, "Human and computer of regional accents and ethnic groups from British English speech," *Computer Speech & Language*, vol. 27, no. 1, pp. 59–74, 2013.

[25] F. Richardson, W. Campbell, and P. Torres-Carrasquillo, "Discriminative N-Gram Selection for Dialect Recognition," in *Interspeech, Brighton, UK*, 2009, pp. 297–307.

[26] A. Lazaridis, J. P. Goldman, M. Avanzi, and P. N. Garner, "Syllable-based Regional Swiss French Accent Identification using Prosodic Features," *Nouveaux Cahiers de Linguistique Française*, vol. 31, pp. 297–307, 2014.

[27] S. Ghazali, R. Hamdi, and M. Barkat, "Speech Rhythm Variation in Arabic Dialects," in *Proceedings of Prosody, Aix-en-Provence, France*, 2002, pp. 331–334.

[28] F. Biadsy and J. Hirschberg, "Using Prosody and Phonotactics in Arabic Dialect Identification," in *Interspeech, Brighton, UK*, 2009, pp. 208–211.

[29] M. Barkat, J. Ohala, and F. Pellegrino, "Prosody as a Distinctive Feature for the Discrimination of Arabic Dialects," in *Eurospeech, Budapest, Hungary*, 1999, pp. 395–398.

[30] R. Hamdi, S. Ghazali, and M. Barkat, "Syllable Structure in Spoken Arabic : a comparative investigation," in *Interspeech*, Lisbonne, Portugal, 2005, pp. 2245–2248.

[31] R. Hamdi, "La variation rythmique dans les dialectes arabes," Ph.D. dissertation, University Lumière Lyon 2 and University 7 Novembre, Tunisie, 2007.

[32] M. Yeou, M. Embarki, and S. Al-Maqtari, "Contrastive focus and F0 patterns in three Arabic dialects," *Nouveaux Cahiers de Linguistique Française*, vol. 28, pp. 317–326, 2007.

[33] S. Ghazali, R. Hamdi, and K. Knis, "Intonational and Rhythmic patterns across the Arabic dialect continuum," in *19th Annual Symposium on Arabic Linguistics, Illinois, USA*, 2005.

[34] D. Crystal, Ed., *A Dictionary of Linguistics and Phonetics*, ser. Lecture Notes in Statistics. Malden: Blackwell, 2008, vol. 6.

[35] P. Boersma and D. Weenink. (2014) Praat v. 5.3.47. Cernak. [Online]. Available: http://www.fon.hum.uva.nl/praat

[36] P. Mertens, "The Prosogram : Semi-Automatic Transcription of Prosody based on a Tonal Perception Model," in *Speech Prosody, Nara, Japan*, 2004, pp. 23–26.

[37] F. Ramus, M. Nespor, and J. Mehler, "Correlates of Linguistic Rhythm in the Speech Signal," *Cognition*, vol. 75, no. 1, pp. 265–292, 1999.

[38] E. Grabe and E. L. Low, "Durational Variability in Speech and the Rhythm Class Hypothesis," *Papers in laboratory phonology*, vol. 7, pp. 515–546, 2002.

[39] V. Dellwo, "Rhythm and Speech Rate: a Variation Coefficient for deltaC," in *Language and language-processing*, 2006, pp. 231–241.

[40] M. Alsulaiman, G. Muhammad, M. A. Bencherif, A. Mahmood, and Z. Ali, "KSU Rich Arabic Speech Database," *Journal of Information*, vol. 16, no. 6, pp. 4231–4253, 2013.

[41] S. Bougrine, "ASLIM: Algerian Arabic Speech Corpus," Laboratoire d'Informatique et de Mathématiques, Laghouat, Algeria, Tech. Rep. RR 06/2015, Jan. 2015.

[42] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.